INVENTEURS DU MONDE NUMÉRIQUE

NOTE D'ANALYSE

Le 1er février 2017

Titres électroniques sécurisés : la centralisation des données biométriques est-elle vraiment inévitable ?

Analyse comparative de quelques architectures

Claude Castelluccia, Daniel Le Métayer

1. Contexte

Le décret du 28 octobre 2016 autorisant la création d'un fichier centralisé de « titres électroniques sécurisés » (TES) a suscité un certain nombre d'interrogations et d'inquiétudes. L'objectif principal mis en avant par le gouvernement est la lutte contre la fraude aux titres d'identité. Cependant, le texte du décret autorise aussi certains accès à la base de données par les agents de la police nationale, de la gendarmerie nationale et des renseignements. De nombreuses voix se sont élevées pour pointer les risques qu'un tel fichier centralisé pourrait représenter en matière de libertés individuelles, et notamment d'atteintes à la vie privée des citoyens. De son côté, le gouvernement avance que ce décret ne fait qu'étendre aux cartes d'identité les dispositions déjà en vigueur pour les passeports et que la solution technique retenue ne permettra pas l'utilisation de la base de données à des fins d'identification[1].

Le présent rapport a été réalisé en parallèle et indépendamment de l'audit conduit par l'ANSSI et la DINSIC[2]. Contrairement à ce dernier, il ne vise pas à analyser la sécurité du système mis en œuvre (auquel les auteurs n'ont pas accès) mais à élargir le débat par l'analyse d'architectures et de solutions alternatives. Son objectif est également d'apporter un éclairage complémentaire sur la protection des données personnelles.

Pour pouvoir se prononcer sur les avantages et inconvénients d'un système de gestion de titres électroniques, il nous paraît nécessaire de :

[1] Cette affirmation est contredite par l'ANSSI et la DINSIC qui constatent dans leur rapport d'audit[2] que « *le système TES peut techniquement être détourné à des fins d'identification* ». Nous expliquons dans la partie 5 pourquoi l'identification est toujours possible dans les systèmes d'authentification reposant sur des liens unidirectionnels.

[2] « Audit du système « Titres Electroniques Sécurisés », Agence nationale de la sécurité des systèmes d'information, Direction interministérielle du numérique et du système d'information et de communication de l'Etat, 13 janvier 2017 : http://www.interieur.gouv.fr/Actualites/Le-systeme-des-titres-electroniques-securises/Systeme-TES-publication-du-rapport-de-l-ANSSI-et-de-la-DINSIC

1. Définir clairement les fonctionnalités souhaitées et les avantages qu'on peut en attendre, notamment par rapport à la situation actuelle et à d'autres solutions.
2. Décrire la solution technique retenue de manière suffisamment précise pour permettre son analyse.
3. Analyser rigoureusement les risques d'atteinte à la vie privée au regard des bénéfices attendus.

Il va de soi qu'un tel document n'a aucune vocation à l'exhaustivité ni à proposer des analyses ou solutions définitives, notre intention étant essentiellement de fournir un cadre pour aborder ces questions de manière rigoureuse. Le jugement que chacun pourra ensuite former sur le bien-fondé de la constitution d'une telle base de données ne repose évidemment pas exclusivement sur ces éléments techniques.

Nous revenons dans la partie 2 sur les incertitudes entourant le projet TES avant d'analyser dans les parties suivantes différentes hypothèses concernant les trois éléments évoqués plus haut: la partie 3 est consacrée aux fonctionnalités du système, la partie 4 décrit différentes solutions techniques et la partie 5 fournit un aperçu de la méthode qui peut être adoptée pour analyser les risques d'atteinte à la vie privée. En conclusion, nous soulignons l'importance des procédures de contrôle (« *accountability* ») en la matière. Nous insistons également sur le fait que le renforcement des moyens de lutte contre la fraude (et la criminalité de manière plus générale) et l'exigence de protection de la vie privée ne sont pas forcément antinomiques.

2. Les incertitudes entourant le nouveau fichier TES

La première source d'inquiétude concernant le nouveau fichier TES est liée à la constitution d'une base de données centralisée regroupant des données sur des dizaines de millions de français[3], qui a pu rappeler d'autres projets contestés comme le « fichier des gens honnêtes » abandonné en 2012 ou encore SAFARI[4] qui a conduit à la création de la CNIL en 1978. D'autres facteurs, de natures variées, ont aggravé la suspicion, comme l'adoption du projet par décret, donc sans débat préalable, alors qu'un pays proche comme la Grande-Bretagne, par exemple, a fini par abandonner un tel projet en 2010 suite à une vague de protestation. Par ailleurs, le texte du décret n'est pas dépourvu d'ambiguïtés et il a suscité des déclarations parfois contradictoires qui ont contribué à alimenter les doutes. Les incertitudes portent aussi bien sur les fonctionnalités attendues que sur la solution technique envisagée.

Objectifs et solutions à préciser

Pour ce qui concerne les fonctionnalités, l'objectif affiché par le gouvernement est de « *simplifier les démarches des usagers et de fiabiliser les titres d'identité en luttant plus efficacement contre la fraude* »[5]. Par ailleurs, le garde des Sceaux précise que le fichier

[3] Tous les détenteurs de carte d'identité ou de passeport. Pour être précis, le ministère de l'Intérieur a décidé, « *afin de répondre aux interrogations qui se sont fait jour* », de « *conditionner le versement dans TES des empreintes digitales des usagers au consentement de ces derniers* » (http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises). L'effet réel de cet assouplissement reste toutefois à vérifier en pratique. En effet, les limites du consentement en matière de protection de la vie privée ont déjà été mises en évidence par différentes études (voir par exemple : « Le consentement au traitement des données personnelles. Perspective comparative sur l'autonomie du sujet », C. Lazaro, D. Le Métayer, Revue Juridique Themis, vol. 48, n° 3, 2015, pp. 765-815).

[4] http://section-ldh-toulon.net/a-l-origine-de-la-Cnil-Safari-ou.html

[5] Site du ministère de l'Intérieur : http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises

TES « *n'a pas vocation à être utilisé à des fins judiciaires* » en ajoutant néanmoins « *sauf dans le cas où les données feraient l'objet de réquisitions* »[6].

A la question de l'apport du nouveau traitement par rapport à la situation actuelle, le principal objectif mis en avant est l'amélioration de l'efficacité de la lutte contre la fraude aux titres d'identité. Par ailleurs, le ministère de l'Intérieur[7] évoque l'offre de « *nouveaux télé-services, comparables à ceux qui existent aujourd'hui pour les passeports : pré-demande en ligne, paiement du timbre dématérialisé en ligne, renouvellement des titres plus rapide, amélioration de la sécurité des CNI, et donc sécurisation de l'identité.* ».

Cependant, la nécessité de constituer une base de données biométriques centralisée pour offrir ces nouveaux services ne paraît pas manifeste. Par ailleurs, la pertinence d'une telle solution en matière de lutte contre la fraude mérite aussi d'être analysée de plus près en regard des modes opératoires utilisés par les fraudeurs[8] (falsification de documents, usage frauduleux d'un document authentique, obtention indue à partir de faux justificatifs, etc.).

S'agissant de la solution technique envisagée, à notre connaissance aucune information précise n'est disponible publiquement. Le site du le ministère de l'Intérieur[9] fait état d'un fichier comportant trois « compartiments » relatifs respectivement à :

- « *des données alphanumériques (l'adresse, le numéro de la demande, le nom du demandeur…) qui figurent sur le formulaire de demande de titre ou CERFA, qui est strictement inchangé* » ;
- « *la photo et aux deux empreintes digitales numérisées* » ;
- « *aux pièces justificatives* ».

Le rapport d'audit de l'ANSSI et de la DINSIC décrit quant à lui un système répartissant les dossiers en deux compartiments : un compartiment alphanumérique et un compartiment de données biométriques dans lequel les « *différentes catégories de données biométriques sont gérées de manière indépendante dans l'optique d'assurer leur cloisonnement*. »

Protection contre l'identification

Le ministère affirme que, s'il est « *possible de remonter au deuxième compartiment, biométrique, à partir des données propres à la demande du titre, l'inverse est impossible. On ne peut accéder à l'identité à partir des données biométriques.* » Cette impossibilité serait non seulement juridique (le décret l'interdit), mais aussi technique. Le fichier TES offrirait donc des fonctions *d'authentification* (« *vérification que la personne qui demande un titre est bien celle qu'elle prétend être au vu du contrôle de conformité des données biométriques que permet la base* ») mais serait mis en œuvre de manière à empêcher toute fonctionnalité *d'identification* (découverte de l'identité d'une personne à partir de données biométriques). Les explications sur cette mise en œuvre fournies par le ministère, notamment dans sa réponse au Conseil national du numérique[10], évoquent une conservation des données biométriques dans une base distincte et séparée de celle des demandes de titres, un lien « *asymétrique* » entre ces bases, et un blocage

[6] Déclaration de Jean-Jacques Urvoas sur son compte Facebook : https://fr-fr.facebook.com/JJ.Urvoas/
[7] Site du ministère de l'Intérieur : http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises
[8] Éléments de connaissance sur la fraude aux documents et à l'identité en 2014, ONDRP : https://www.inhesj.fr/sites/default/files/fichiers_site/ondrp_ra-2015/fraude_documents_cr.pdf
[9] http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises
[10] http://www.interieur.gouv.fr/Actualites/Communiques/Fichier-TES-Courrier-de-Bernard-Cazeneuve-au-President-du-Conseil-national-du-numerique

technique « *garanti par une cryptographie spécifique et un lien unidirectionnel* ». Le rapport d'audit de l'ANSSI et la DINSIC affirme cependant que « *le système TES peut techniquement être détourné à des fins d'identification* » et recommande la prise en compte des « préconisations du Référentiel Général de Sécurité concernant les mécanismes cryptographiques mis en œuvre pour construire les liens unidirectionnels. » Les éléments disponibles publiquement sont trop vagues pour permettre une véritable analyse technique. Cependant, certains scientifiques sont sceptiques[11] sur la possibilité même de l'existence d'une solution technique offrant la fonctionnalité d'authentification tout en interdisant celle d'identification. Cette question importante est discutée dans la partie 5 de ce document.

Sécurisation de la base de données

Pour ce qui est de la sécurisation de la base de données, le ministère mentionne que « *des outils cryptographiques sont mis en œuvre pour les données biométriques. De même, les pièces justificatives sont cryptées. Des barrières physiques que l'on dénomme HSM ou pare-feux sont également déployées et le système TES bénéficie d'une bulle sécurisée et de serveurs dédiés. Il faut par ailleurs préciser que le réseau sur lequel l'application centrale est opérée n'est pas sur Internet mais interne au ministère de l'Intérieur. Il s'agit donc d'une application qui est conservée à distance solide des réseaux publics, comme l'est la base TES depuis 2008* » et signale que « *le système TES et plus généralement les applications hébergées à distance des réseaux publics au sein du ministère de l'Intérieur, n'ont fait l'objet d'aucun hacking ces dernières années.* », en concluant par : « *Les faits parlent d'eux-mêmes.* »[12]

Même si les mesures évoquées paraissent de bon sens, un examen plus approfondi serait nécessaire pour apporter des assurances plus solides[13]. L'audit réalisé par l'ANSSI et la DINSIC avait pour objet de répondre à ce besoin mais ses conclusions ne sont pas entièrement rassurantes puisque le rapport comprend onze recommandations portant sur des aspects critiques du système comme la robustesse du lien unidirectionnel, le cloisonnement, le traçage des requêtes, la sécurité des serveurs (« non conformes à l'état de l'art ») et la gouvernance du système. Par ailleurs, il serait nécessaire également d'analyser l'équilibre entre les risques inhérents à toute solution centralisée et les bénéfices qu'elle peut apporter par rapport à des architectures distribuées.

La question du consentement

Pour tenter de répondre aux inquiétudes soulevées par le projet, le gouvernement a indiqué que « *dans le cadre d'une demande ou d'un renouvellement d'une carte nationale d'identité, le versement des empreintes digitales du demandeur du titre seront soumis à son consentement express et éclairé. Ainsi, le recueil des empreintes reste obligatoire. Mais le refus du versement des empreintes dans la base centralisée TES n'empêchera pas la délivrance du titre. En revanche, ce versement simplifie et facilite l'émission d'un nouveau titre et permet de lutter efficacement contre l'usurpation d'identité : les Français qui y renoncent renonceront également aux services associés.* »

---

[11] Dont les auteurs de ce rapport. Voir aussi : http://www.lsv.ens-cachan.fr

[12] http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises

[13] La question n'est pas de savoir si la sécurité du système n'a jamais été compromise mais de fournir des garanties que les risques ont bien été étudiés et que des mesures satisfaisantes ont été prises pour y remédier. On sait que les risques de fuites de données ne sont jamais nuls et qu'ils peuvent devenir massifs quand les bases de données sont centralisées. A titre de rappel, l'Office of Personnel Management (OPM) a révélé en 2015 la fuite de 5,6 millions d'empreintes digitales de fonctionnaires américains : http://www.lemondeinformatique.fr/actualites/lire-5-6m-d-empreintes-digitales-de-fonctionnaires-americains-derobees-62451.html

Cependant, on connaît de manière générale les limites du consentement en matière de collecte des données personnelles[14] et on peut se demander si, dans un tel contexte et face aux arguments sécuritaires des agents chargés de l'émission des titres, un grand nombre de citoyens seront suffisamment informés ou auront suffisamment de détermination pour refuser ce consentement.

Contrôle des accès aux données personnelles

Pour ce qui concerne le contrôle des accès aux données personnelles, le site du ministère de l'Intérieur met en avant un avantage du nouveau système par rapport à la situation existante : « *A cet égard, on soulignera que les données biométriques propres à la CNI existent déjà, sont déjà relevées par les agents de mairie et consultées par les agents de préfecture (photo et empreintes). Simplement, elles sont conservées sous format papier. L'inconvénient majeur du format papier, c'est qu'il rend complexe la traçabilité des consultations auxquelles il donne lieu. Alors qu'en versant dès aujourd'hui les CNI dans l'application TES, nous sommes désormais en mesure de garantir aux Français la traçabilité parfaite des consultations de leurs données biométriques, notamment avec un système d'horodatage.* »

Cependant, le traçage ne représente qu'une condition nécessaire du contrôle : il serait utile, pour rendre l'argument convaincant, de préciser les modes effectifs de contrôle, son organisation, les entités impliquées et les garanties qui peuvent être fournies aux citoyens quant à l'existence d'audits véritablement indépendants. Cette question importante, de ce qu'on appelle en anglais l'« accountability » (responsabilité, au sens d'obligation de « rendre compte »), est discutée en conclusion.

3. Définition des fonctionnalités

Le décret du 28 octobre 2016 s'inscrit dans le cadre du plan PPNG (« Plan Préfectures Nouvelles Générations »). L'objectif affiché par le gouvernement est double[15] : « *Il s'agit à la fois de simplifier les démarches des usagers et de fiabiliser les titres d'identité en luttant plus efficacement contre la fraude* ». Nous nous focalisons dans ce document sur la finalité de lutte contre la fraude qui est généralement mise en avant pour justifier la constitution d'une base centralisée de données biométriques.

Nous pouvons distinguer trois phases principales dans la gestion des titres d'identité[16] :

1. La phase *d'émission du premier titre* durant laquelle une personne demande son premier titre d'identité. L'identification de la personne est généralement effectuée en utilisant des certificats administratifs, tels que des actes de naissance. Durant cette phase, des empreintes et des photos du demandeur sont collectées et enregistrées. L'enregistrement peut être effectué (de manière non exclusive) dans une base de données ou sur un support physique individuel sécurisé (carte à puce).

2. La phase de *renouvellement d'un titre* durant laquelle une personne demande le renouvellement de son titre d'identité, son premier titre n'étant plus valide ou ayant été perdu ou volé. L'identification du demandeur peut être effectuée en

[14] "Le consentement au traitement des données personnelles. Perspective comparative sur l'autonomie du sujet", C. Lazaro, D. Le Métayer, Revue Juridique Themis, vol. 48, n° 3, 2015, pp. 765-815, http://www.academia.edu/12524523/_Le_consentement_au_traitement_des_données_personnelles._Perspective_comparative_sur_lautonomie_du_sujet

[15] Site du ministère de l'Intérieur : http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises

[16] https://www.service-public.fr/particuliers/vosdroits/F21089

utilisant l'ancien titre et des justificatifs sous forme de documents « papier » (seule possibilité en cas de non présentation du premier titre), tels que des actes de naissance. On pourra comparer les empreintes et photos du demandeur avec celles qui sont stockées dans la base de données et/ou sur la carte à puce, selon les options d'architectures retenues.

3. La phase *d'utilisation du titre* durant laquelle une personne doit décliner son identité. L'identification de la personne est souvent effectuée par la présentation d'une carte nationale d'identité ou d'un passeport (sous forme papier ou électronique).

Les deux fonctionnalités suivantes sont particulièrement utiles pour lutter contre la fraude :

1. *La vérification de l'identité (VI)* : vérification qu'une identité est valide (elle n'est pas fausse) et qu'elle correspond à la personne qui se présente (elle n'est pas usurpée).
2. *La détection de doublons (DD)* : détection des doublons d'identité (tentative d'usurpation d'identité), c'est à dire qu'une empreinte, ou un trait biométrique, a été enregistré avec différents noms.

Même si la fonctionnalité DD peut être utilisée pendant la phase d'utilisation d'un titre, elle est surtout importante pendant les phases d'émission et de renouvellement des titres. Par ailleurs, elle s'applique de la même manière dans ces deux phases. Par conséquent, et par souci de simplification, nous ne la considérons que pour la phase d'émission du premier titre dans la suite du document.

4. Définition des architectures

Différents choix d'architecture et de techniques sont possibles pour mettre en œuvre un système de gestion de titres d'identité. Chaque solution peut permettre d'atteindre de manière plus ou moins satisfaisante les fonctionnalités attendues (ou un sous-ensemble de ces fonctionnalités) et présenter des risques plus ou moins importants en matière de vie privée. Nous décrivons dans cette partie quelques options architecturales avant d'analyser la manière dont elles permettent d'atteindre les fonctionnalités de vérification d'identité et de détection de doublons. Les risques qu'elles peuvent présenter en matière de vie privée sont étudiés dans la partie suivante. L'objectif de ce document est de fournir un aperçu de la diversité des options possibles et de montrer la nécessité de procéder de manière méthodique pour les comparer. Il n'est donc pas question, ici, d'analyser ou de citer toutes les architectures possibles. Nous nous focalisons sur les options suivantes :

- **Architecture A1** : cette architecture repose sur la mise en place de deux fichiers centralisés, constitués et gérés par l'administration. Le premier fichier contient, pour chaque citoyen enregistré[17] : les données d'état civil (nom, prénoms, date et lieu de naissance, sexe et données relatives à la filiation), certaines données personnelles additionnelles (couleur des yeux, taille, adresse) et une image de la signature du demandeur. Le deuxième fichier contient les données biométriques de chaque citoyen, c'est à dire des représentations numérisées du visage et des empreintes digitales. Les données biométriques sont chiffrées et liées aux données d'état civil par des liens unidirectionnels (ou asymétriques). Il est ainsi possible de retrouver simplement les données biométriques à partir des données

[17] http://www.interieur.gouv.fr/Actualites/Le-fichier-des-titres-electroniques-securises

d'état civil d'une personne (authentification), par exemple en suivant un simple lien (ou pointeur) mais il est plus difficile[18] de retrouver les données d'état civil à partir des données biométriques car le lien (ou pointeur) inverse n'existe pas. Cette architecture semble correspondre à la solution préconisée par le gouvernement pour le nouveau système TES.

- **Architecture A2 :** dans cette architecture, les titres électroniques (cartes d'identité, passeports) sont équipés d'une carte à puce qui contient les données personnelles et les données biométriques de leur détenteur. Le fichier centralisé ne contient que les données d'état civil (nom, prénoms, date et lieu de naissance, sexe et données relatives à la filiation) et certaines données personnelles additionnelles (couleur des yeux, taille, adresse). Les données biométriques sont donc stockées uniquement sur une carte à puce contrôlée par le détenteur du titre.

- **Architecture A3** : la troisième option peut être vue comme une combinaison des architectures A1 et A2 précédentes. Elle comporte un fichier biométrique centralisé et des titres électroniques équipés d'une carte à puce stockant les données d'état civil et biométriques du détenteur. Cette solution semble correspondre à l'architecture adoptée pour la gestion des passeports électroniques français (système TES existant).

- **Architecture A4** : la quatrième option est similaire à A3 à la différence près que le fichier centralisé des données d'état civil ne comporte aucun lien vers les données biométriques. Les deux fichiers sont donc complètement séparés et il est impossible d'associer (dans un sens ou dans l'autre) une donnée biométrique à des données d'état civil.

Le tableau 1 ci-dessous résume les architectures considérées dans ce document ainsi que la solution actuelle (A0) qui ne comporte ni fichier biométrique centralisé ni carte à puce.

	Architecture avec fichiers biométriques centralisés	**Architecture sans fichier biométrique centralisé**
Titre avec carte à puce	**A3** (solution TES passeport) **A4**	**A2**
Titre sans carte à puce	**A1** (solution TES cartes d'identité)	**A0** (solution actuelle)

Tableau 1 : architectures considérées

La première question qui se pose à propos des architectures ci-dessus est celle de leurs capacités à assurer les fonctionnalités décrites dans la partie précédente. Le tableau 2 compare, pour les quatre architectures considérées, le niveau de protection (en matière de lutte contre la fraude) apporté pour chacune de ces fonctionnalités. Nous distinguons les niveaux « fort » quand la protection repose sur des moyens de sécurisation forts (par exemple à partir des données biométriques), « faible » quand elle est apportée uniquement par des moyens visuels et « impossible » quand elle ne peut pas être apportée.

[18] Nous préférons utiliser l'expression « plus difficile » plutôt que « impossible » qui nous paraît un objectif inatteignable. Cette question est discutée dans la partie 5.

Nous calculons également, pour chaque architecture considérée, à titre d'illustration, un « score de fonctionnalité ». Ce score est calculé en additionnant les trois sous-scores des trois phases de la gestion des titres d'identité (émission, renouvellement, utilisation). La valeur 2 correspond à une fonctionnalité garantie « fortement », la valeur 1 à une fonctionnalité garantie « faiblement » et la valeur 0 à fonctionnalité n'est pas fournie (« impossible »). Le score final est donc une valeur entre 0 et 6, où 0 indique qu'aucune fonctionnalité souhaitée n'est fournie alors qu'un score de 6 indique que toutes les fonctionnalités sont garanties « fortement ». Ces scores sont présentés dans le tableau 2 (sous le nom de l'architecture considéré).

L'analyse du tableau 2 montre que :

- L'un des maillons faibles des quatre architectures présentées est la phase de vérification de l'identité (VI) lors de *l'émission du premier titre*. Cette phase est très importante car l'authenticité des données versées dans la base, et par conséquent la sécurité des phases suivantes, en dépend directement. Dans toutes les architectures, la sécurité de cette phase est faible car elle repose essentiellement sur des justificatifs qui doivent être fournis par le demandeur et qui sont falsifiables. Il conviendrait de renforcer la sécurité de cette phase en replaçant, par exemple, les justificatifs « papier » par des justificatifs électroniques signés par les entités émettrices ou en généralisant l'usage de la solution 2D-Doc de sécurisation des justificatifs développée par l'ANTS [19] . Cependant, la sécurisation de justificatifs comme les certificats de naissance n'est pas un problème simple à résoudre, car il ne suffit pas d'authentifier l'émetteur de certificat (ce qui peut être fait, par exemple, via une signature électronique) : il faut aussi s'assurer que le certificat appartient bien au demandeur du titre. Cette question est délicate et mériterait une étude plus détaillée.

- Une solution reposant sur les cartes à puce, sans faire appel à une base de données centralisée (A2), ne permet pas de détecter les doublons d'identité lors d'une demande de renouvellement d'un titre. Cependant, les protections apportées lors de la phase de renouvellement dépendent directement de la fiabilité de la phase d'émission du premier titre. En effet, si la phase d'émission du premier titre est sûre, et les titres d'identité eux-mêmes peuvent être considérés comme infalsifiables (les cartes à puce étant des dispositifs sécurisés), l'usurpation d'identité est alors très difficile et la phase de détection des doublons devient accessoire, voire inutile.

- Un titre d'identité sans carte à puce ne permet une vérification forte de l'identité lors de son utilisation que dans les scénarios permettant l'accès à une base de données centrale. Or, l'accès à cette base de données devra être strictement limité et contrôlé pour des raisons évidentes de sécurité[20]. Dans les autres cas, cette vérification restera essentiellement visuelle à partir d'un titre papier, facilement falsifiable, et fournira donc un niveau de protection faible. Les solutions qui font appel à une carte à puce ont l'avantage considérable de fournir une vérification d'un niveau de sécurité très élevé (les cartes à puce étant très difficilement falsifiables).

- L'architecture A4 permet d'assurer les mêmes fonctionnalités que l'architecture A3 tout en garantissant l'indépendance totale (absence de liens) entre le fichier des données biométriques et le fichier des données d'état civil. Cette

[19] Voir le site de l'ANTS (Agence Nationale des Titres Sécurisés) : https://ants.gouv.fr/Les-solutions/2D-Doc
[20] La réponse du ministre de l'Intérieur au président du CNNum suggère le déploiement d'un réseau dédié et d'une application « conservée à distance solide des réseaux publics ».

caractéristique de l'architecture A4 permet, comme nous le montrons dans la section 5, de réduire les risques d'atteinte à la vie privée tout en préservant un niveau de fonctionnalité équivalent.

	Premier Titre	**Renouvellement Titre**	**Utilisation Titre**
A0 **Score :2.5/6**	**VI** (**faible**) : vérification réalisée sur les justificatifs « papier ». **DD** (**impossible**) : vérification impossible car aucune base centralisée.	**VI** (**faible**) : vérification réalisée en comparant l'empreinte du demandeur avec celle qui apparaît sur l'ancien titre[21] (et vérification de justificatifs « papier »).	**VI** (**faible**) : vérification réalisée en comparant l'empreinte ou la photo du demandeur avec celle qui apparaît sur le titre.
A1 **Score :5/6**	**VI** (**faible**) : vérification réalisée sur les justificatifs « papier ». **DD** (**fort**) : vérification réalisée en vérifiant si l'empreinte du demandeur existe déjà dans la base biométrique.	**VI** (**fort**) : vérification réalisée en comparant l'empreinte du demandeur avec celle qui correspond à son identité dans la base (et vérification de justificatifs « papier »).	**VI** (**faible** ou **fort**) : vérification automatique impossible localement, car le titre ne possède aucune puce, mais possible en interrogeant la base (si le vérifieur peut y accéder).
A2 **Score : 4.5/6**	**VI** (**faible**) : vérification réalisée sur les justificatifs « papier ». **DD** (**impossible**) : vérification impossible car aucune base de données centralisée	**VI** (**fort**) **:** vérification réalisée en comparant le nom et l'empreinte du demandeur aux empreintes dans la carte à puce[22] (et vérification de justificatifs « papier »).	**VI** (**fort**) : vérification de l'identité possible en comparant l'empreinte de l'utilisateur avec l'empreinte stockée dans la carte à puce.
A3 **Score : 5.5/6**	**VI** (**faible**) : vérification réalisée sur les justificatifs « papier ». **DD** (**fort**) : vérification réalisée en vérifiant si l'empreinte du demandeur existe déjà dans la base biométrique.	**VI** (**fort**) : vérification réalisée en comparant le nom et l'empreinte du demandeur aux empreintes dans la carte à puce[23] (et vérification de justificatifs « papier »).	**VI** (**fort**) : vérification de l'identité possible en comparant empreinte de l'utilisateur avec l'empreinte stockée dans la carte à puce. Vérification également possible en interrogeant la base (si le vérifieur peut y accéder).
A4 **Score :5.5/6**	**VI** (**faible**) : vérification réalisée sur les justificatifs « papier ». **DD** (**fort**) : vérification réalisée en vérifiant si l'empreinte du demandeur existe déjà dans la base biométrique.	**VI** (**fort**) : vérification réalisée en comparant le nom et l'empreinte du demandeur aux empreintes dans la carte à puce[24] (et vérification de justificatifs « papier »).	**VI** (**fort**) : vérification de l'identité possible en comparant empreinte de l'utilisateur avec l'empreinte stockée dans la carte à puce.

Tableau 2 : protection contre la fraude

[21] Si le demandeur a perdu sa carte, identique à la phase « premier titre ».
[22] Ibid.
[23] Ibid.
[24] Ibid.

5. Analyse des risques d'atteinte à la vie privée

La mise en place d'un système de gestion des titres sécurisés pouvant présenter des risques d'atteinte à la vie privée, il nous paraît nécessaire de procéder au préalable à une étude d'impact (« Privacy Impact Assessment » ou « PIA » en anglais).

Etude d'impact

Une telle étude doit prendre en compte une multitude de facteurs qui peuvent avoir une incidence sur les risques. Il convient de considérer notamment les types de données personnelles traitées, les différents intervenants, les sources de risques[25], les évènements redoutés[26], les attaques[27] possibles, et enfin leurs impacts potentiels sur les personnes. L'analyse des risques est d'autant plus complexe qu'il est généralement nécessaire d'envisager un très grand nombre de scenarios correspondant à des combinaisons de paramètres variés. La réalisation d'une véritable étude d'impact sort clairement du cadre de ce document et notre objectif dans cette partie est essentiellement d'esquisser la démarche à suivre.

S'agissant d'un système de gestion de titres d'identité électroniques, une étude d'impact en matière de vie privée devrait permettre d'évaluer au moins les deux types de risques majeurs associés aux traitements biométriques, que sont :

1. Les risques liés à l'usage des données personnelles en mode *identification* par le gouvernement.
2. Les risques résultant de fuites ou de vols de données personnelles commis par des sources externes (cybercriminels, gouvernements étrangers, etc.) ou internes (employés ou sous-traitants mal intentionnés, etc.).

On considère généralement deux composantes essentielles pour évaluer les risques : la gravité de leurs impacts sur les personnes et leur vraisemblance[28]. La gravité des impacts des deux risques évoqués ci-dessus est majeure. Ces impacts peuvent être d'ordre physique (suicide suite à un vol d'identité ou à la publication de données personnelles), matériel (pillage d'un compte bancaire suite à une usurpation d'identité) ou psychologique (sensation de surveillance, de harcèlement, etc.).
La vraisemblance d'un risque dépend directement de l'architecture envisagée et des contre-mesures mises en place. Dans le reste de cette partie, nous analysons de manière succincte la vraisemblance des deux risques cités plus haut pour les cinq architectures décrites dans la partie précédente.

Risques d'identification

L'utilisation du système à des fins d'identification par le gouvernement consisterait à retrouver l'identité d'une personne à partir d'une empreinte ou d'une photo. Il est intéressant de noter que seules les solutions A2 et A4 offrent une protection satisfaisante à cet égard :

[25] Cybercriminels, gouvernement, états étrangers, activistes, employés indélicats, fraudeurs, etc.
[26] Identification, surveillance, usurpation d'identité, spams, phishing, etc.
[27] Intrusion, vol, complicité, falsification de documents, etc.
[28] https://www.cnil.fr/sites/default/files/typo/document/CNIL-PIA-1-Methode.pdf

- **A1 et A3** : Une protection contre l'identification est introduite dans A1 et A3 par l'utilisation de liens unidirectionnels, rendant plus difficile le passage d'une empreinte aux données d'état civil correspondantes. Cependant, cette protection demeure très faible car il suffirait d'interroger la base de données avec les noms des personnes susceptibles d'en faire partie (par exemple tous les citoyens français) pour reconstituer la base complète avec les liens bidirectionnels[29]. Il paraît difficile, voire impossible, de se protéger *techniquement* contre un tel risque à partir du moment où *toutes les données sont contrôlées par une seule entité*. L'introduction de liens unidirectionnels complique l'identification, mais ne l'empêche pas de façon absolue. De même, le fait de ne stocker qu'un gabarit ou un condensat des empreintes ou des photos, comme il est parfois proposé[30], ne constitue qu'une faible protection contre ce risque, car il suffirait de comparer les condensats au lieu des empreintes afin de retrouver l'identité de la personne en question[31]. Par ailleurs, même sans reconstituer la base, il est possible de l'interroger pour vérifier certaines identités. Il est aisé, par exemple lors d'une manifestation, d'effectuer une recherche à partir d'une liste de noms de « suspects » potentiels (opposants, syndicalistes, etc.).

- **A2** protège contre ce type de risque car elle ne stocke aucune information biométrique dans la base centralisée**.** Pour ce qui est de A4, bien que cette architecture comporte un fichier biométrique, elle ne permet pas l'identification car ce fichier ne comporte aucun lien (ni dans un sens ni dans l'autre) avec le fichier de données d'état civil. Il est donc impossible d'identifier les données biométriques.

Risques liés aux fuites de données

Les données stockées sur une plateforme de gestion des identités sont très sensibles et doivent être protégées avec la plus grande attention. On sait de longue date en sécurité informatique que la centralisation représente un facteur de risque majeur car elle désigne à un attaquant une cible très tentante. Une fuite de données peut résulter d'une attaque externe (intrusion), d'une attaque interne (employé malveillant) ou encore d'une négligence (mauvais choix de configuration, sauvegardes non sécurisées, etc.). Par ailleurs, même dans les cas où la probabilité d'une telle fuite pourrait être considéré comme très faible, il convient de considérer l'impact majeur qu'elle pourrait avoir en matière de vie privée[32]. En effet de telle données pourraient, par exemple, être utilisées par des fraudeurs (cybercriminels) pour réaliser des usurpations d'identité *à grande échelle*[33] ou pour identifier des personnes à partir de leurs photos ou empreintes. De toutes les architectures considérées précédemment, seules les architectures A2 et A4 protègent véritablement contre ce type de risques : A2 parce qu'elle ne comporte pas de base centralisée et A4 parce qu'elle ne permet pas d'établir de lien entre données biométriques et données d'état civil.

Le tableau 3 ci-dessous résume les éléments de discussion concernant la vraisemblance des risques considérés ici. Nous avons utilisé la gradation suivante pour quantifier ces risques : possible (2), difficile (1) et impossible (0). Nous

[29] Ce qu'on appelle parfois une attaque de « force brute ».
[30] Voir par exemple : Symmetric Hash Functions for Fingerprint Minutiae, S. Tulyakov, F. Farooq, V. Govindaraju, *Third International Conference on Advances in Pattern Recognition*, ICAPR 2005, Springer
[31] Lire http://www.pellegrini.cc/2016/11/la-biometrie-des-honnetes-gens-reloaded/
[32] Le risque est parfois quantifié en multipliant la probabilité d'un l'évènement redouté par ses impacts potentiels.
[33] Même si le nombre d'usurpations d'identité est limité, chaque fraude peut avoir un impact potentiellement très grand pour les victimes. L'introduction d'un fichier biométrique centralisé peut donc conduire à remplacer des évènements (fraudes) peu probables et ayant une incidence sur un nombre limité de personnes en évènements très peu probables mais faisant un très grand nombre de victimes (attaques à grande échelle).

présentons également pour chaque architecture un « score de risque ». Ce score est calculé en additionnant les sous-scores de chaque source de risques (Etat français, sources internes, source externes). Le score final est donc une valeur entre 0 et 6, 0 indiquant que les risques sont négligeables et 6 que les risques sont très élevés. Même s'il ne fournit évidemment qu'une appréciation très grossière du niveau de risque, un score global de ce genre permet de faire ressortir des éléments de comparaison utiles pour l'aide à la décision.

Risques / Archi-tectures	**Etat français (surveillance)**	**Sources internes (employés, sous-traitants, etc.)**	**Sources externes (cybercriminels, états tiers, etc.)**
A0[34] **Score : 2.5/6**	Identification : **impossible**	Accès non-légitime aux données d'état civil et biométriques : **difficile**	Fuite de données d'état civil : **difficile** Fuite de données biométriques : **difficile** Identification des données biométriques : **difficile**
A1 Score : 5/6	Identification : **possible**	Accès non-légitime aux données d'état civil et biométriques : **possible**[35]	Fuite de données d'état civil : **difficile**[36] Fuite de données biométriques : **difficile** Identification des données biométriques : **difficile**
A2 Score : 1.3/6	Identification : **impossible**	Accès non-légitime aux données d'état civil : **possible** Accès non-légitime aux données biométriques : **impossible**	Fuite de données d'état civil : **difficile** Fuite de données biométriques : **impossible** Identification des données biométriques : **impossible**
A3 Score : 5/6	Identification : **possible**	Accès non-légitime aux données d'état civil et biométriques : **possible**	Fuite de données d'état civil : **difficile** Fuite de données biométriques : **difficile** Identification des données biométriques : **difficile**
A4 Score : 1.66/6	Identification : **impossible**	Accès non-légitime aux données d'état civil : **possible** Accès non-légitime aux données biométriques : **impossible**	Fuite de données d'état civil : **difficile** Fuite de données d'état civil et biométriques : **difficile** Identification des données biométriques : **impossible**

Tableau 3 : Etude préliminaire de la vraisemblance des risques en matière de vie privée

[34] La nature des risques de l'architecture A0 est différente de celle des autres architectures. En effet, A0 n'utilise pas de fichier centralisé, les risques existent, mais sont « locaux » et d'ampleur moins importante.
[35] Même si les données biométriques sont chiffrées, des sources de risques internes (administrateur système malveillant ou soudoyé, sous-traitant indélicat, etc.) peuvent avoir accès aux clefs de chiffrement et donc aux données chiffrées. Ce risque peut être limité par des mesures de dissuasion (fichiers logs sécurisés, permettant de garder une trace des accès, et contrôlés par des auditeurs indépendants).
[36] Les bases de données d'état civil et biométriques étant chiffrées, il est difficile d'accéder aux données en cas de fuite.

Eléments de comparaison

La comparaison des tableaux 2 et 3, et la figure 1 ci-dessous, qui présente les scores de risques en fonction des scores de fonctionnalité, montrent clairement que la constitution d'une base de données biométriques centralisée et « identifiante » (A1 et A3, soit les architectures TES pour les cartes d'identité et les passeports) représente une source de risque majeure d'atteinte à la vie privée au regard de la solution actuelle (A0) et d'autres solutions techniques envisageables (A2 et A4).

L'analyse esquissée dans ce document montre également que les architectures reposant sur des cartes à puce, notamment A4 qui possède un score de fonctionnalité élevé tout en minimisant les risques, sont des options de mise en œuvre qui méritent d'être étudiées.

Il va de soi que cette analyse de risques doit elle-même être développée de manière plus approfondie à partir d'éléments d'informations plus complets, notamment sur les architectures, avant d'en tirer des conclusions plus affirmées. Par ailleurs, nous n'avons pas évoqué ici d'autres solutions techniques plus sophistiquées comme celles qui sont listées dans l'avis du Conseil National du numérique[37] ou proposées par des industriels et des équipes de recherche[38]. De même, nous nous sommes concentrés sur les risques d'atteinte à la vie privée, sans considérer les conséquences indirectes pour un Etat comme la France, aussi bien en termes d'image, de crédibilité que de confiance, d'une fuite massive de données biométriques.

Enfin, ce rapport s'est focalisé sur l'objectif principal affiché par le gouvernement, à savoir la fiabilisation de l'émission des titres d'identité pour lutter contre la fraude. Il ne considère pas l'utilisation des données par les forces de l'ordre ou les officiers de police judiciaire. Par exemple, la justice doit pouvoir, dans le cadre de réquisitions, demander à avoir accès à l'ensemble des données (données civiles et données biométriques. Cette fonctionnalité nécessiterait une analyse spécifique. Il conviendrait en particulier, comme il est suggéré dans le rapport d'audit de l'ANSSI et la DINSIC, d'étudier la dissociation des deux systèmes. Cette option, proche dans l'esprit de l'architecture A4 du présent rapport[39], permettrait d'adopter une démarche de type « privacy-by-design » minimisant, pour chaque scenario, les données collectées et leur utilisation.

[37] Avis du CNNum sur le fichier TES, 12 décembre 2016, https://cnnumerique.fr

[38] Voir par exemple : la technologie "Match-on-Card": http://www.morpho.com/en/match-card , les travaux sur les condensats de données biométriques : Symmetric Hash Functions for Fingerprint Minutiae (http://dl.acm.org/citation.cfm?id=2094169); Secure Method for Biometric-Based Recognition with Integrated Cryptographic Functions (https://www.hindawi.com/journals/bmri/2013/623815/), Privacy-Preserving Biometric Database (http://www.cs.haifa.ac.il/~orrd/crypt/biometric.pdf) ou encore la proposition de François Pellegrini: http://www.pellegrini.cc/2016/11/la-biometrie-des-honnetes-gens/

[39] On pourrait imaginer, par exemple, d'étendre l'architecture A4 avec une table de correspondance entre les deux fichiers (état civil et données biométriques). Conformément à la recommandation 1 du rapport de l'ANSSI, cette table pourrait être doublement chiffrée par le ministère et par une entité tierce de telle façon que ni l'un ni l'autre ne puisse la déchiffrer seul.

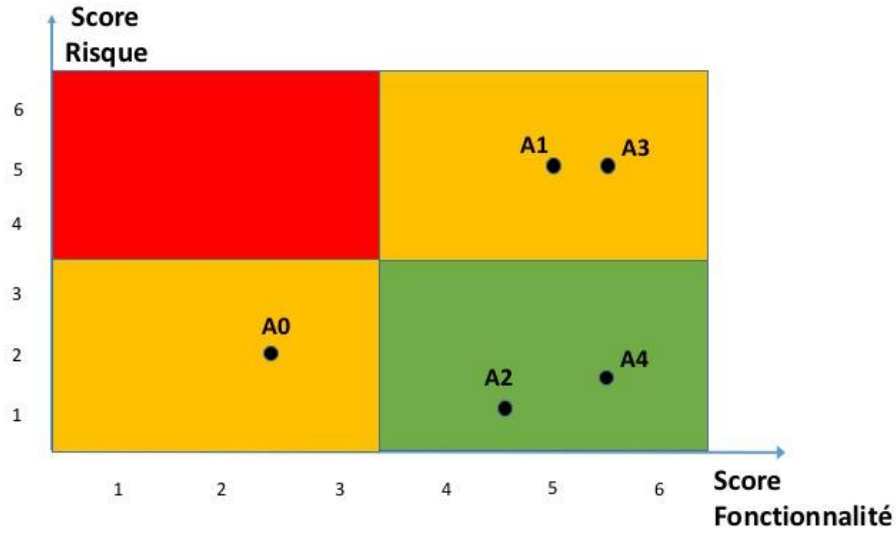

Figure 1 : risques vs fonctionnalités

6. Conclusion : des précautions préalables aux contrôles a posteriori

Indépendamment du récent décret autorisant la création du fichier TES, l'objectif du présent document est d'alerter les pouvoirs publics et l'opinion sur la nécessité d'adopter une démarche plus rigoureuse et de prendre toutes les précautions nécessaires avant de déployer des systèmes pouvant présenter des risques majeurs en matière de vie privée[40]. Cette méthode n'est pas nouvelle : elle est consacrée par le nouveau règlement européen sur les données personnelles[41] et elle a fait l'objet de travaux scientifiques[42]. L'analyse des risques en est la pierre angulaire et les résultats de cette analyse doivent permettre d'éclairer le débat, notamment pour ce qui concerne l'architecture du système, et de les justifier les options retenues. Nous avons également illustré dans ce document le fait que différentes options sont possibles – et bien d'autres mériteraient d'être analysées [43] – avec des conséquences variables aussi bien en matière de protection contre la fraude que de risques d'atteinte à la vie privée.

En conclusion, il convient aussi d'insister sur la nécessité de mettre en place des mesures de responsabilisation (« accountability ») et des contrôles rigoureux pour minimiser les risques d'usages détournés des données personnelles, notamment de dévoiement de la base de données à des fins de surveillance. Des mesures techniques (système auditable, avec des logs sécurisés permettant de tracer toutes les opérations) doivent rendre possible ces contrôles[44] mais elles sont insuffisantes si elles ne sont pas accompagnées de mesures organisationnelles tout aussi fiables : il est primordial notamment d'assurer que les audits pourront être conduits par des experts indépendants. Cet ensemble de garanties techniques, organisationnelles et juridiques est une condition *sine qua non* du rétablissement d'une certaine confiance des citoyens envers ce type de systèmes et ceux qui les mettent en place.

---

[40] Cette démarche mériterait d'ailleurs d'être également adoptée pour les systèmes déjà déployés, notamment pour le système TES des passeports qui comporte plusieurs dizaines de millions d'entrées.
[41] Journal officiel de l'Union européenne, 4 mai 2016, Règlement général sur la protection des données (article 35), http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AL%3A2016%3A119%3ATOC
[42] Voir notamment : S. J. De, D. Le Métayer, Privacy Risk Analysis, Morgan & Claypool Publishers, septembre 2016, S. J. De, D. Le Métayer, A Risk-Based Approach to Privacy by Design (extended version), Rapport de Recherche Inria 9001, décembre 2016.
[43] Voir notamment l'annexe de l'avis du CNNum sur le fichier TES, https://cnnumerique.fr
[44] L'ANSSI et la DINSIC insistent également sur cette exigence de traçabilité dans leur rapport d'audit.

Remerciements

Les auteurs tiennent à remercier Hervé Chabanne, Mathieu Cunche, Antoine Petit et Vincent Roca pour leurs remarques constructives sur des versions antérieures de ce document.

A propos d'Inria

Inria, institut national de recherche dédié au numérique, promeut « l'excellence scientifique au service du transfert technologique et de la société ». Inria emploie 2600 collaborateurs issus des meilleures universités mondiales, qui relèvent les défis des sciences informatiques et mathématiques. Son modèle ouvert et agile lui permet d'explorer des voies originales avec ses partenaires industriels et académiques. Inria répond ainsi efficacement aux enjeux pluridisciplinaires et applicatifs de la transition numérique. Inria est à l'origine de nombreuses innovations créatrices de valeur et d'emplois